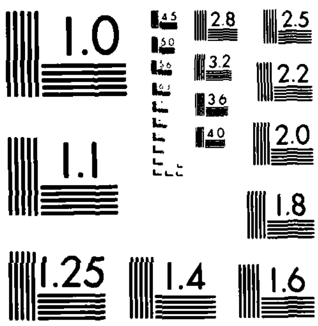
AD-A149 337 SPACEFILLING CURVES AND THE PLANER TRAVELLING SALESMAN PROBLEM REVISION(U) GEORGIA INST OF TECH ATLANTA SCHOOL OF INDUSTRIAL AND SYSTEMS.. L K PLATZMAN ET AL. OCT 84 1/1

UNCLASSIFIED N00014-80-K-0709 F/G 12/2 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

GEORGIA INSTITUTE
OF TECHNOLOGY
ATLANTA, GEORGIA 30332

84 12 05 018

SPACEFILLING CURVES AND THE PLANAR TRAVELLING SALESMAN PROBLEM

Loren K. Platzman and John J. Bartholdi, III

School of Industrial and Systems Engineering

Georgia Institute of Technology

Atlanta, Georgia 30332

Abstract

This paper analyzes the performance of a novel heuristic to obtain the minimal-length tour of N given points in the plane: they are sequenced as they appear along a spacefilling curve. The algorithm consists essentially of sorting, so it is easily coded and requires only O(N) memory and O(N log N) operations. Its performance is shown to be competitive with that of other available methods.

Key words: Planar travelling salesman problem, heuristic algorithm, spacefilling curve.

The authors have been supported by the Office of Naval Research under Contract No. N00014-80-k-0709, and by National Science Foundation under Grant ECS-8351313.

Originally PDRC 83-02, February, 1983. Revised October, 1984. Submitted to JACM.

1. Introduction. The travelling salesman problem (TSP) is to construct a circuit of minimum total length that visits each of N given points. Even in the plane, this problem is NP-complete [10]. Thus, instances of practical interest cannot be solved *exactly* in reasonable time. Accordingly, attention has focussed on fast algorithms that generate good but not necessarily optimal tours.

The authors have recently introduced a practical approach of appealing simplicity to this problem [1-3]. It is exceptionally well suited to manual execution (routes may be generated by nontechnical personnel without a computer and even, after an initial setup, without a map [2]), and consequently, it has been adopted by a variety of commercial, charitable and public organizations to generate daily delivery routes. It is based on a spacefilling curve ψ, a continuous mapping from the unit interval $C = [0,1]$ *onto* the unit square $S = [0,1]^2$, and is performed as follows:

SPACEFILLING HEURISTIC.

1) For each point $p \in S$ to be visited, compute a $\theta \in C$ such that $p = \psi(\theta)$.

2) Sort the points by their corresponding θ's.

In other words, this heuristic visits points in sequence of their appearance along the spacefilling curve.

Our work was inspired by Karp [8], who introduced a family of O(N log N) algorithms to construct tours of length arbritrarily close to optimal. (The *effort* grows rapidly as optimality is approached, however!) Karp's algorithms divide S into rectangles sufficiently small so that each contains a given number, say t, of points. A routing problem is solved exactly within each rectangle, and the subtours are patched together. Our heuristic may be viewed as a limiting case of Karp's algorithms: the square is subdivided into subsquares that each contain but one point, and the patching procedure for joining subrectangles is predetermined and specified by the spacefilling curve.

The spacefilling heuristic is of special interest because it is based on spacefilling curves. Originally devised as topological counterexamples nearly a century ago, these were long regarded as "mathematical monstrosities". It is only recently that their usefulness has been recognized [11]. Our work represents the first application of spacefilling curves to combinatorial optimization.

The spacefilling heuristic is appealing due to its ease of execution. But it is necessary to show that it also performs well, and moreover, is competitive with other methods. Standard combinatorial arguments are inappropriate because they rely on the combinatorial structure of the problem and spacefilling curves, by their very nature, eliminate this structure. Our analysis utilizes properties of measure-preserving transformations, metric spaces, and convexity.

This paper establishes worst-case bounds on the heuristic tour length (Theorem 3) and the ratio of heuristic to optimal tour lengths (Theorem 4), and almost sure bounds (for increasingly large random point sets) on the heuristic tour length (Theorem 5.3) and the length of the longest link along the heuristic tour (Theorem 5.1). To streamline the presentation, we provide the analysis for a specific curve in the plane; following [3], our methods can be generalized to the TSP in d-space, and to more general combinatorial problems, such as matching and clustering.

Table 1 summarizes our performance analysis of the spacefilling heuristic. Also included are the performances of comparable methods cited by Bentley [5] as particularly simple. These are:

Nearest Neighbor (NN). Start at an arbitrary point and successively visit the nearest unvisited point. After all points have been visited, return to the start.

Minimum Spanning Tree (MST). Construct the minimum spanning tree of the point set and duplicate all the links of the tree. Sequence the points as they would appear in a traversal of the doubled tree. Pass through the sequence and remove all representations after the first of each point.

Strip. Partition the square into $\sqrt{N}$ vertical strips. Visit the points in each strip in order (alternately top-to-bottom and bottom-to-top) and visit the strips from left to right. Return to the starting point.

We know of no rigorous statistical study of the expected tour lengths for the comparison heuristics, but our informal tests indicate practically identical behavior among these algorithms for large problems consisting of uniformly distributed points in the square.

	NN	MST	Strip	Spacefilling
Memory	O(N)	O(N)	O(N)	O(N)
Worst-case effort				
To solve	$O(N^2)^*$	$O(N^2)^*$	O(N log N)	O(N log N)
To modify	Re-solve	Re-solve*	O(log N)	O(log N)
Worst-case ratio				
Bound	O(log N)	2	O(N)	O(log N)
Known	$O\left(\frac{\log N}{\log\log N}\right)$	2	O(N)	4.7
Longest tour	$2.15\sqrt{N}$	$3.04\sqrt{N}$	$2\sqrt{N}+1+\sqrt{2}$	$2\sqrt{N}$
Performance on nonuniform data	Good	Good	Poor	Good
Ease of coding	Good	Poor	Good	Good

Table 1. *Comparison of simple TSP heuristics. To make as clean a comparison as possible, we have considered only the most straightforward implementations and have omitted enhancements such as sophisticated data structures or subroutines designed to mitigate pathological behavior. However, efficient sorting procedures have been assumed where appropriate. An asterisk indicates that the entry may be reduced, but only at considerable programming expense.*

Across a spectrum of criteria, the spacefilling heuristic is comparable to or better than other commonly considered heuristics for the TSP. Unlike these heuristics, however, the spacefilling heuristic may be modified in the spirit of Karp to produce tours arbitrarily close to optimal, as follows:

ARBITRARILY CLOSE SPACEFILLING HEURISTIC.

1) Perform the spacefilling heuristic, and let $p_1, \ldots, p_N$ represent the points sequenced according to the heuristic tour.

2) For $i=1,\ t+1,\ 2t+1,\ \ldots,$ determine the shortest path starting at p_i, passing through $p_{i+1}, \ldots, p_{i+t-2}$ in any sequence, and ending at p_{i+t-1}: adjust the heuristic tour accordingly.

This algorithm requires $O(N \log N + N\, 2^t)$ effort. The analytical techniques in this paper and in [4] and [8] can be applied to show that it produces tours approaching optimal length as $N >> t$ and $t \to \infty$.

Halton and Terada [7] have devised a partitioning heuristic for which they make impressive performance claims. The arbitrarily close spacefilling heuristic can match this performance by solving a sequence of N-point problems, $N=1,2,\ldots,$ with t a function of N such as $t = \log \log \log \log N$. The sort may be performed in asymptotically linear time (*e.g., by BINSORT*). The tours are asymptotically optimal, yet the effort is almost linear, that is, $O(N \log \log \log N)$ a.s. Of course, the benefits of asymptotic optimality will not become evident until $N >> 2^{2^{2^{2^2}}} \sim 10^{10^4}$.

Fundamental properties of the particular spacefilling curve used in this study are given in Section 2. Various performance bounds are derived in the next three sections.

2. Fundamental properties of the spacefilling curve.

The particular spacefilling curve upon which we base our analysis is the limit of the sequence of curves shown in Figure 1. The θ required in step 1 of the spacefilling heuristic is evaluated according to the function THETA, given in Appendix A.

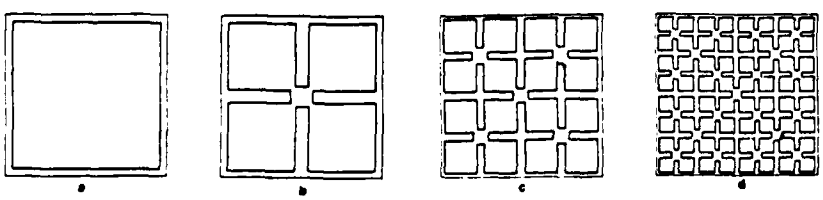


Figure 1. *Successive approximations to the spacefilling curve ψ.*

We first consider the computational effort required to perform the spacefilling heuristic when THETA is used to implement step 1. If the arguments X and Y of THETA are integer multiples of 2^{-k} (that is, if they are given to k binary digits), then THETA will call itself k times. Therefore evaluating THETA (step 1 of the heuristic) requires an effort that depends on k but not N; it consists of bit sampling and shifting, and may be arranged to require O(k) bit operations. Furthermore, the value returned by THETA will be an integer multiple of 2^{-2k-2} (that is, it will be given to 2k+2 digits). So each comparison of θ values in the sort (step 2 of the heuristic) requires an effort O(k) that does not depend on N. Consequently we have:

PROPOSITION 2.1. *The spacefilling heuristic requires* O(N log N) *effort.*

REMARK. If a sorting procedure such as BINSORT [9,13] is used, then the heuristic may be performed in linear expected time.

PROPOSITION 2.2. *The spacefilling heuristic requires O(N) memory.*

Since the heuristic tour is simply a sorted list, every subsequence of the heuristic tour is itself a heuristic tour. Thus, if the set of points to be visited changes slightly, the current solution need be modified only locally to produce a new heuristic tour. This observation has important practical consequences in applications where routes must be updated frequently [2]. It is formally stated as follows:

PROPOSITION 2.3. *Inserting a point into or deleting a point from an N-point heuristic tour requires O(log N) effort.*

Next we establish three fundamental properties of the curve which will be required in subsequent sections. The first two are evident from Figure 1.

LEMMA 2.4. *For any integers* $k > 0$ *and* $0 \leq i < 2^k$, *the set*

$$\{\psi(\theta) \mid i\ 2^{-k} \leq \theta \leq (i+1)\ 2^{-k}\}$$

is an isosceles right triangle whose right angle lies at $\psi((i+0.5)2^{-k})$.

LEMMA 2.5. *The mapping* ψ *is measure preserving. That is, for any interval* I *in* C,

$$\mathrm{area}\{\psi(I)\} = \mathrm{length}\{I\}.$$

Lemma 2.5 is important throughout Section 5, where we consider random points uniformly distributed on S. These points are unlikely to have finite binary representations, and an infinite-precision version of THETA must be imagined to generate the θ's to which they correspond. Since ψ is measure preserving, these θ's will be uniformly distributed on C and almost surely uniquely determined (since the set of points to which many θ's correspond has measure zero).

The final property of ψ to be considered here expresses the notion that a spacefilling curve preserves nearness: points close together in C map (via ψ) onto points close together in S. We take the measure of "nearness" on the square S to be Euclidean distance, denoted by $D[\cdot,\cdot]$. As is evident in Figure 1, we can view C as a circuit since $\psi(0) = \psi(1)$; thus the natural metric on C is

$$\Delta[\theta,\theta'] = \min\{|\theta-\theta'|,\ 1-|\theta-\theta'|\}.$$

The following lemma is implicit in [11, p. 65] although no proof is given.

LEMMA 2.6. *For any* $\theta,\theta' \in C$,

$$D^2[\psi(\theta),\ \psi(\theta')] \leq 4\ \Delta[\theta,\ \theta'].$$

PROOF. Assume without loss of generality that $0 < \theta'-\theta < 0.5$. Let $\theta_{i,k} = \max(\theta,\ \min(\theta',\ i\ 2^{-k}))$ and

$$Q_k = \sum_{i=0}^{2^k-1} D^2[\psi(\theta_{i,k})\ ,\ \psi(\theta_{i+1,k})].$$

First suppose C1: $\theta' \geq i\,2^{-k}$ and $\theta \leq (i+1)\,2^{-k}$. Then by Lemma 2.4, $\theta_{i,k}$ and $\theta_{i+1,k}$ both lie within an isosceles right triangle whose right angle lies at $\hat{\theta}_{i,k} = (i+0.5)\,2^{-k}$. By Lemma 2.5, the area of this triangle is 2^{-k}. Since any distance within a right triangle cannot exceed the length of the hypotenuse,

$$D^2[\psi(\theta_{i,k}),\ \psi(\theta_{i+1,k})] \leq 4 \cdot 2^{-k}. \qquad (2.1)$$

But if C1 does not hold, then $\theta_{i,k} = \theta_{i+1,k}$, which also implies (2.1). So (2.1) holds for all i,k.

Assume further C2: $\theta < \hat{\theta}_{i,k} < \theta'$. The Pythagorean Theorem (cosine law) yields

$$\begin{aligned} &D^2[\psi(\theta_{i,k}),\ \psi(\theta_{i+1,k})] \\ &\quad \leq D^2[\psi(\theta_{i,k}),\ \psi(\hat{\theta}_{i,k})] + D^2[\psi(\hat{\theta}_{i,k}),\ \psi(\theta_{i+1,k})] \\ &\quad = D^2[\psi(\theta_{2i,k+1}),\ \psi(\theta_{2i+1,k+1})] + D^2[\psi(\theta_{2i+1,k+1}),\ \psi(\theta_{2i+2,k+1})]. \qquad (2.2) \end{aligned}$$

But if C2 does not hold, then $\theta = \theta_{2i,k+1} = \theta_{2i+1,k+1}$ or $\theta' = \theta_{2i+1,k+1} = \theta_{2i+2,k+1}$, which also implies (2.2). So (2.2) holds for all i,k.

Clearly, $Q_0 = D^2[\psi(\theta),\ \psi(\theta')]$. By (2.2) and the definition of Q_k,

$$Q_0 \leq Q_1 \leq \ldots \leq Q_k. \qquad (2.3)$$

And by (2.1) and the definition of Q_k,

$$Q_k \leq (\lceil 2^k\theta' \rceil - \lfloor 2^k\theta \rfloor) \cdot 4 \cdot 2^{-k} \leq 4\,(\theta'-\theta) + 8 \cdot 2^{-k}. \qquad (2.4)$$

The limit as $k \to \infty$ of (2.3)-(2.4) yields the asserted inequality. ■

3. Worst-case heuristic tour length.

We now show that the spacefilling heuristic cannot produce a very long tour.

THEOREM 3. *The heuristic tour length cannot exceed $2\sqrt{N}$.*

PROOF: Let $\theta_1,\ldots,\theta_N$ be the sorted list generated by the spacefilling heuristic, and set $\Delta_i=\theta_{i+1}-\theta_i$, $i=1,\ldots,N-1$, and $\Delta_N=1+\theta_1-\theta_N$. By Lemma 2.6,

$$\text{Heuristic tour length} \le \sum_{i=1}^{N} 2\sqrt{\Delta_i}\ . \tag{3.1}$$

But $\Delta_i\ge 0$ and $\sum_{i=1}^{N}\Delta_i = 1$. So the bound (3.1), a symmetric concave function of $\{\Delta_i\}$, achieves its maximum at $\Delta_i\equiv 1/N$. ■

4. Worst-case ratio of heuristic to optimal tour lengths.

Although Theorem 3 guarantees that the heuristic tour cannot be very long, the optimal tour could be considerably shorter. The worst-case instance we have found produces a heuristic tour that is 4.707 times longer than optimal, and we conjecture that the heuristic tour length will never exceed the optimal tour length by more than a constant factor. But the strongest result we have proved is the following.

THEOREM 4. $\dfrac{\textit{Heuristic tour length}}{\textit{Optimal tour length}} = O(\log N).$

PROOF. Let Π be the set of N points (in S) to be visited. If $\lambda_1,\ldots,\lambda_N$ denote the N link lengths along the heuristic tour, and $H(t) = \#\{k \mid \lambda_k>t\}$ (where $\#\{\cdot\}$ denotes cardinality), then the heuristic tour length L may be written

$$L = \sum \lambda_k = \sum \int_0^\infty 1\{\lambda_k>t\}\,dt = \int_0^\infty H(t)\,dt, \tag{4.1}$$

where $1\{\cdot\}$ denotes the indicator function. To establish the claim, we will derive upper bounds on H(t).

The principal bound is derived from "Minkowski's sausage" [11], denoted by $T(\epsilon)$, and defined to be the set of points (in the plane) that lie within ϵ of at least one point in the locus of the optimal tour (the union of points along the N segments that form the tour). By a simple geometric argument, there are $c,c'>0$ such that

$$\text{area}\{T(\epsilon)\} \le c\epsilon L^* + c'\epsilon^2, \qquad \text{for all } \Pi \text{ and all } \epsilon>0. \tag{4.2}$$

where L^* is the length of the optimal tour through Π. (The first term of (4.2) represents asymptotic behavior as $\epsilon \to 0$ and $T(\epsilon)$ becomes a ribbon of length L^* and width 2ϵ; the second term represents asymptotic behavior as $\epsilon \to \infty$ and $T(\epsilon)$ becomes a circle of radius ϵ.)

Now for arbitrary $t>0$, let $m = \lceil 2/t \rceil$. Partition C into disjoint intervals $I_1, \ldots, I_{m^2}$, each of length m^{-2}. By Lemma 2.6, the points in $\psi(I_k)$ lie within $2/m$ of each other, so

$$\psi(I_k) \cap \Pi \text{ is nonempty} \Rightarrow \psi(I_k) \subseteq \Pi(2/m), \qquad (4.3)$$

where $\Pi(\epsilon)$ denotes the set of points (in the plane) that lie within ϵ of at least one point in Π. If the distance between any two consecutive points along the heuristic tour exceeds $2/m$, then the θ's corresponding to these points cannot lie in the same interval, so

$$H(t) \le \#\{k \mid \lambda_k > 2/m\} \le \#\{k \mid \psi(I_k) \cap \Pi \text{ is nonempty}\}. \qquad (4.4)$$

By Lemma 2.5, $\text{area}\{\psi(I_k)\} = m^{-2}$. Thus (4.3)-(4.4) yield

$$H(t) \le m^2 \, \text{area}\{\Pi(2/m)\}. \qquad (4.5)$$

Clearly $\Pi(\epsilon) \subseteq T(\epsilon)$. So (4.2), (4.5), and [illegible] definition of m become

$$H(t) \le 2cmL^* + 4c' \le 4cL^*/t + 2cL^* + 4c'. \qquad (4.6)$$

Another bound on $H(t)$ can be established by noting that the distance between any two points in Π cannot exceed L^* or $\sqrt{2}$, so

$$H(t) = 0, \qquad t > \min\{L^*, \sqrt{2}\}. \qquad (4.7)$$

Furthermore, since there are only N links in the heuristic tour,

$$H(t) \le N. \qquad (4.8)$$

Now combining (4.1) and (4.6)-(4.8) gives

$$L \le \int_0^{\min(L^*, \sqrt{2})} \min\{N, 4cL^*/t + 2cL^* + 4c'\} \, dt$$

$$\le \int_0^{L^*} \min\{N, 4cL^*/t\} \, dt + \int_0^{\sqrt{2}} 2cL^* \, dt + \int_0^{L^*} 4c' \, dt.$$

Evaluating the integrals yields $L/L^* = O(\log N)$. ■

5. Stochastic analysis. Let $\{p_i\}$ be an infinite sequence of independent uniformly distributed points in S, and let

L_N = length of the heuristic tour through $\{p_1,\ldots,p_N\}$,

L_N^* = length of the optimal tour through $\{p_1,\ldots,p_N\}$.

In a classic work, Beardwood, Halton and Hammersley [4] showed that $L_N^* / \sqrt{N} \rightarrow \beta^*$ a.s. (The constant β^* has been experimentally determined to be 0.765.) The purpose of this section is to produce similar asymptotic bounds on the heuristic tour lengths L_N. We also examine the length of the longest link along the heuristic tour, and show that it grows only slightly faster than the average link length.

To begin, we note that the *nicest possible convergence*, $L_N / \sqrt{N} \rightarrow \beta$, does not hold. We prove instead a result whose practical implications are the same, that $L_N / \sqrt{N}$ "converges" to a narrow interval $[\beta^-,\beta^+]$ in the sense that

$$\beta^- \le \liminf L_N/\sqrt{N} \quad \text{and} \quad \limsup L_N/\sqrt{N} \le \beta^+, \quad \text{a.s.} \tag{5.1}$$

Since the optimal tour is no longer than the heuristic tour, $\beta^* \le \beta^-$, and by Theorem 3, $\beta^+ \le 2$, so β^- and β^+ exist. This section will establish tight bounds on these constants. *Numerical evaluation* of the bounds shows that β^- and β^+ lie within a range no greater than 0.956 ± 0.001. Thus, for large N, the heuristic tour length will be about 25% above optimal.

Our analysis may be modified, as in [4], so it applies to independent points nonuniformly distributed in the plane. If the points have density $f(x,y)$, and $K(f) = \int\int \sqrt{f(x,y)}\, dy\, dx$, then the optimal tour grows as $K(f)\, \beta^* \sqrt{N}$ and the heuristic tour grows as between $K(f)\, \beta^- \sqrt{N}$ and $K(f)\, \beta^+ \sqrt{N}$. Thus the heuristic tour remains about 25% longer than the optimal tour. If the points are uniformly distributed over any region of area $A>0$, then $K(f)= \sqrt{A}$.

This has a useful consequence: it implies that the N/K-th point in a tour will lie at 1/K-th the tour length from the start of the tour, when N is large. (To see this, observe that N/K consecutive θ's span 1/K-th the length of C, and constitute independent, uniformly distributed points over the image under ψ of that range. Since ψ is measure preserving, that image has area 1/K.) An important application is the formation of delivery routes for K vehicles by partitioning the travelling salesman tour into K segments, each to be travelled by a single vehicle [2]. All subtours contain equal numbers of points, but it is desirable that they be equally long. Partitioning the spacefilling heuristic tour produces routes which tend to be of equal length. This is not true of the optimal tour, or of tours formed by other heuristic methods.

Even the largest distance between two consecutive points along the heuristic tour cannot greatly exceed the average interpoint distance, as the following theorem demonstrates. Thus, the spacefilling heuristic produces a tour whose performance is good with respect to the "bottleneck" (maximum link length) criterion [6] as well as the customary total tour length criterion.

THEOREM 5.1. *Let* E_N *denote the length of the longest link in the heuristic tour of* $\{p_1, \ldots, p_N\}$. *Then, for all* $a>1$,

$$\lim_{N\to\infty} \mathrm{Prob}\left(E_N \geq 2\,a\,\sqrt{(\ln N)/N}\right) = 0.$$

PROOF: Let e_N be the largest distance between consecutive θ's in the sorted list produced by the spacefilling heuristic. In light of Lemma 2.6, we need only show that

$$\lim_{N\to\infty} P\left(e_N \geq a^2\,(\ln N)/N\right) = 0.$$

The distance between θ's is not affected by linear shifts within C. Subtracting the smallest θ from each θ in the list produces a θ at 0 (equivalently at 1) and a sorted list of N-1 independent uniformly distributed θ's between 0 and 1. These points determine N intervals of which the largest has length e_N.

We now construct a new random variable whose distribution coincides with that of e_N. Let $\{t_i\}$ be an infinite sequence of independent exponentially distributed random variables of mean 1, and let $S_N = \sum_{i=1}^{N} t_i$. That is, S_N is the time of the N-th arrival of a unit intensity Poisson process. Also let $M_N = \max_{i=1,\ldots,N} t_i$. If $S_N = T$, then $S_1,\ldots,S_{N-1}$ will be conditionally distributed as a sc ted list of N-1 independent uniform random variables over [0,T]. Thus, M_N/T is conditionally distributed as e_N, given $S_N = T$. Integrating over T, this becomes: M_N/S_N is distributed as e_N. It remains to show that

$$\lim_{N\to\infty} P\{M_N/S_N \geq a^2 (\ln N)/N\} = 0.$$

Clearly,

$$P\{M_N/S_N \geq a^2 (\ln N)/N\}$$

$$= P\{M_N/\ln N \geq a^2 S_N/N\}$$

$$\leq P\{M_N/\ln N \geq a \text{ or } S_N/N \leq 1/a\}$$

$$\leq P\{M_N/\ln N \geq a\} + P\{S_N/N \leq 1/a\}.$$

Since M_N is the maximum of N independent exponential random variables,

$$P(M_N/\ln N \geq a) = 1 - (1 - N^{-a})^N \leq N^{1-a},$$

where we have used the inequality $(1-x)^N \geq 1 - Nx$, $0<x<1$. This expression vanishes as $N\to\infty$. And, by the Law of Large Numbers, $P\{S_N/N \leq 1/a\}$ vanishes as well. ■

REMARK. We have observed that $E_N/\sqrt{(\ln N)/N}$ does not converge as N grows, but that it most often lies between 1.1 and 1.3.

We now return to the bounds β^- and β^+ in (5.1). We first examine the relationship between the *random* heuristic tour lengths L_N and their expectations $E\{L_N\}$.

LEMMA 5.2. *The random sequence* $[L_N^* - E\{L_N^*\}]/\sqrt{N}$ *converges almost surely to zero. Thus (5.1) holds with*

$$\beta^- = \liminf E\{L_N/\sqrt{N}\} \qquad \beta^+ = \limsup E\{L_N/\sqrt{N}\}$$

PROOF. Steele's proof [12] that $[L_N^* - E\{L_N^*\}]/\sqrt{N} \to 0$ a.s. applies to the heuristic tour length as well. ■

Since β^- and β^+ are determined by the *deterministic* sequence of expectations $E\{L_N\}/\sqrt{N}$, the remaining analysis will be concerned solely with expectations. We now show that β^- and β^+ are quite close, so that "for all practical purposes" $L_N/\sqrt{N}$ converges almost surely.

Define

$$m(t) = \int_0^1 D[\psi(\theta), \psi(\theta+t \bmod 1)]\, d\theta., \tag{5.2}$$

to be the expected Euclidean distance between two uniformly distributed points in S whose corresponding θ's lie exactly distance t apart. Also let $g(t) = \sqrt{t}\, e^{-t}$ and $\gamma(t) = \sum_{k=-\infty}^{\infty} 4^k\, t\, g(4^k t)$. Clearly $\gamma(4t) = \gamma(t)$. Furthermore, γ displays very little variation:

$$\begin{aligned} \gamma^- &= \min \gamma(t) \sim 1.275 \\ \gamma^+ &= \max \gamma(t) \sim 1.281. \end{aligned} \tag{5.3}$$

These numbers have the following significance.

THEOREM 5.3. *There is a constant* r^* *such that*

$$\gamma^- r^* \le \beta^- \le \beta^+ \le \gamma^+ r^*$$

and r^* *is arbitrary closely determined by*

$$\left| r^* - \int_a^{4a} t^{-3/2}\, m(t)\, dt \right| \le 6a, \qquad \forall\ 0<a<0.25$$

To establish Theorem 5.3 we require two lemmas whose proofs are given in Appendix B. The first shows that $m(t)$ displays a certain limiting behavior as $t \to 0$.

LEMMA 5.4. *There is a continuous function* r *on* $(0,1]$ *such that*

(a) $\quad \left| m(t) - \left[r(t) \cdot \sqrt{t} \right] \right| \le 2t^{3/2}$.

(b) $\quad r(t/4) = r(t)$.

(c) $\quad 0 \le r(t) \le 2$.

Let Δ_N be a random variable whose distribution is the same as that of the difference between two consecutive θ's in the sorted list of N independent values. Clearly,

$$E\{L_N\}/\sqrt{N} = \sqrt{N}\, E\{\text{link length}\} = \sqrt{N}\, E\{m(\Delta_N)\}. \tag{5.4}$$

For large N, the sorted list of θ's will approximate a Poisson process on C, and so the distribution of Δ_N will approach a negative exponential of mean $1/N$. Moreover, since Δ_N is likely to be small, the limiting form of $m(\cdot)$ (given by Lemma 5.4a) may be substituted into (5.4). These manipulations are summarized by

LEMMA 5.5.

$$\lim_{N\to\infty} \left\{ E\{L_N\}/\sqrt{N} - \sqrt{N} \int_0^\infty r(t)\, \sqrt{t}\, N\, e^{-Nt}\, dt \right\} = 0.$$

PROOF OF THEOREM 5.3. Let $r^* = \int_a^{4a} r(t)/t\, dt$; by Lemma 5.4b, this integral does not depend on a. By Lemma 5.5, $E\{L_N\}/\sqrt{N}$ approaches

$$\begin{aligned}
&\sqrt{N} \int_0^\infty r(t)\, \sqrt{t}\, N\, e^{-Nt}\, dt \\
&= \int_0^\infty r(t)\, g(Nt)\, N\, dt \\
&= \sum_{k=-\infty}^{\infty} \int_{4^k a}^{4^{k+1} a} r(t)\, g(Nt)\, N\, dt \\
&= \int_a^{4a} r(t)\, \gamma(Nt)/t\, dt \qquad \text{(by Lemma 5.4b)}
\end{aligned}$$

and the asserted result follows from Lemmas 5.2 and 5.4a. ■

REMARK. These bounds may be made even tighter by using the stronger result $r(t/2)=r(t)$ instead of Lemma 5.4b. This places β^- and β^+ within 10^{-5} of each other. Our estimates, given at the start of this section, are based on these tighter bounds, as well as Monte-Carlo simulation of (5.2) to compute r^*. The statistical error of our simulation was $\pm\ 10^{-3}$.

Acknowledgements. We thank Steven Hackman, Arnon Rosenthal, Richard Serfozo, Craig Tovey, and John Vande Vate for their helpful comments.

REFERENCES

[1] Bartholdi, J. J. and Platzman, L. K. (1982). An O(n log n) Planar Travelling Salesman Heuristic Based on Spacefilling Curves. *Operations Research Letters* 1 121-125.

[2] Bartholdi, J. J., Platzman, L. K., Collins, R. L., and Warden, W. H. (1983). A Minimal Technology Routing System for Meals-on-Wheels. Interfaces 13 1-8.

[3] Bartholdi, J. J. and Platzman, L. K. (1984). Heuristics Based on Spacefilling Curves for Combinatorial Problems in the Plane. Submitted to *Oper. Res.*

[4] Beardwood, J., Halton, J. H. and Hammersley, J. (1959). The shortest Path Through Many Points. *Proc. Cambridge Philosophical Society* 55 299-327.

[5] Bentley, J. L. (1984). A case study in applied algorithm design. *IEEE Computer* 75-88.

[6] Garfinkle, R. S. and Gilbert, K. C. (1978). The Bottleneck Travelling Salesman Problem: Algorithms and Probabilistic Analysis. *J. ACM* 25 435-449.

[7] Halton, J. H. and Terada, R. (1982). A Fast Algorithm for the Euclidean Travelling Salesman Problem, Optimal with Probability One. *SIAM J. Comput.* 11 28-46.

[8] Karp, R. M. (1977). Probabilistic Analysis of Partitioning Algorithms for the Travelling Salesman Problem in the Plane. *Math. Oper. Res.* 2 209-224.

[9] Knuth, D. (1973). *The Art of Computer Programming, Vol. 3:* Sorting and Searching. Addison-Wesley, Reading, MA.

[10] Papadimitriou, C. H. (1977). The Euclidean Travelling Salesman Problem is NP-complete. *Theoret. Comput. Sci.* 4 237-244.

[11] Mandelbrot, B. B. (1983). *The Fractal Geometry of Nature.* W. H. Freeman, New York.

[12] Steele, J. M. (1981). Complete Convergence of Short Paths and Karp's Algorithm for the TSP. *Math. Oper. Res.* 6 374-378.

[13] Weide, B. W. (1978). Statistical Methods in Algorithm Design and Analysis. Ph.D. Thesis, Carnegie-Mellon University, Pittsburgh, PA.

APPENDIX A

AN ALGORITHM TO PERFORM STEP 1 OF THE SPACEFILLING HEURISTIC

Let:

ABS(A) = A if A≥0, = -A if A<0.
INT(A) = the largest integer not larger than A.
FRACT(A) = A-INT(A).
MIN(A,B) = A if A≤B, = B if A>B.
MOD(A,B) = B*FRACT(A/B).
NV(X,Y) = the 'number' of vertex (X,Y) of the unit square, counting clockwise from the origin, i.e., NV(0,0)=0, NV(0,1)=1, NV(1,1)=2, NV(1,0)=3.

The algorithm is given as a recursive function:

```
FUNCTION THETA(X,Y):

  If X=1 and Y=1 then RETURN(0.5)

  Q=NV(MIN(INT(2*X),1), MIN(INT(2*Y),1)))        (Q identifies the quad-
                                                  rant containing (X,Y))

  T=THETA(2*ABS(X-0.5), 2*ABS(Y-0.5))            (T is the position along the
                                                  subcurve in quadrant Q)

  If MOD(Q,2)=1 then T=1-T                       (Visit the vertices of
                                                  a quadrant clockwise)

  RETURN(FRACT((Q+T)/4 + 7/8))
```

APPENDIX B

PROOFS OF LEMMAS REQUIRED IN THEOREM 5.3.

PROOF OF LEMMA 5.4. Let

$$m^-(t) = \int_0^{1-t} D[\psi(\theta), \psi(\theta+t)]\, d\theta$$

and

$$m^+(t) = m^-(t) + 2\, t^{3/2} = m^-(t) + \int_{1-t}^{1} 2\, \sqrt{t}\, d\theta.$$

Clearly $m^-(t) \le m(t)$, and by Lemma 2.6, $m(t) \le m^+(t)$. Since ψ visits S by visiting four subsquares of S, each identical to S at half the scale (see Figure 1),

$$m^-(t) \le 2\, m^-(t/4)$$

$$m^+(t) \ge 2\, m^+(t/4).$$

So $2^k\, m^-(t/4^k)$ and $2^k\, m^+(t/4^k)$ approach a common limit $m^*(t)$ satisfying

$$m^*(t) = 2\, m^*(t/4) \qquad (B.1)$$

and

$$|m(t)-m^*(t)| \le 2\, t^{3/2}. \qquad (B.2)$$

By Lemma 2.6,

$$m(t) \le 2\, \sqrt{t}, \qquad (B.3)$$

and by the triangle inequality

$$m(t+\epsilon) \le m(t) + m(\epsilon) \le m(t) + 2\, \sqrt{\epsilon},$$

so $m(\cdot)$ is continuous. Now let $r(t) = m^*(t)/\sqrt{t}$. Continuity of r follows from that of m; (a) follows from (B.2); (b) follows from (B.1) and (c) follows from (B.3). ■

PROOF OF LEMMA 5.5. The exact density of Δ_N is

$$f_N(t) = N\,(1-t)^{N-1}, \qquad 0 \le t \le 1,$$

so (5.4) may be written

$$E\{L_N\}/\sqrt{N} = \sqrt{N} \int_0^1 m(t)\, f_N(t)\, dt.$$

Now

$$\left| E\{L_N\}/\sqrt{N} - \sqrt{N} \int_0^1 r(t)\, \sqrt{t}\, f_N(t)\, dt \right|$$

$$\le \quad \sqrt{N} \int_0^1 \left| m(t) - \left[r(t)\cdot\sqrt{t} \right] \right| f_N(t)\, dt$$

$$\le \quad \int_0^1 2\, (Nt)^{3/2}\, (1-t)^{N-1}\, dt \qquad \text{(by Lemma 5.3a)}$$

$$\le \quad \int_0^1 2\, (Nt)^{3/2}\, e^{-(N-1)t}\, dt$$

$$\le \quad \int_0^\infty 2\, (Nt)^{3/2}\, e^{-(N-1)t}\, dt$$

$$= \quad O(1/N).$$

It remains to show that the following sequence vanishes as $N \to \infty$,

$$\left| \left[\sqrt{N} \int_0^1 r(t)\, \sqrt{t}\, f_N(t)\, dt \right] - \left[\sqrt{N} \int_0^1 r(t)\, \sqrt{t}\, N\, e^{-Nt}\, dt \right] \right|$$

$$\le \quad \int_0^1 2\, N^{3/2}\, \sqrt{t}\, \left| (1-t)^{(N-1)} - e^{-Nt} \right| dt \qquad \text{(by Lemma 5.4c)}$$

$$\le \quad \int_0^\infty 2\, \sqrt{\tau}\, \left\{ \left| \max(0, 1-\tau/N)^{(N-1)} - e^{-\tau} \right| \right\} d\tau. \qquad (\tau = Nt)$$

Since $(1-\tau/N)^N$ converges upward to $e^{-\tau}$, the integrand of this upper bound converges pointwise to zero as $N \to \infty$, and by Lebesgue's dominated convergence theorem, the limit integral vanishes as well. ∎

END

FILMED

2-85

DTIC